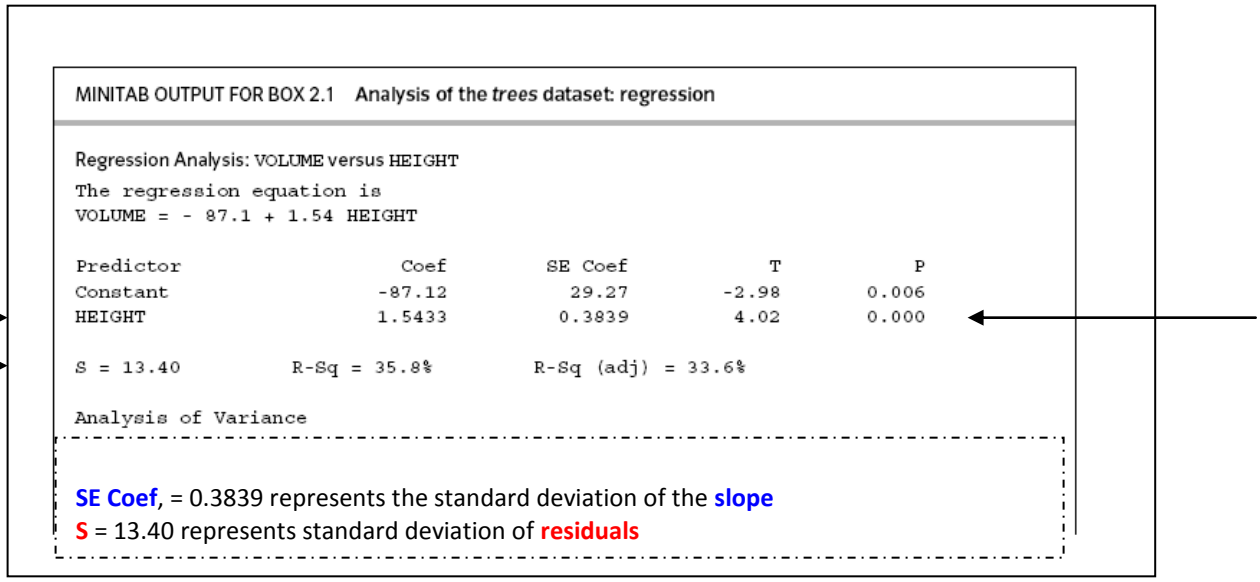
AP Statistics – AP Exam Review **Name:_____**

Linear Regression & Computer Output: Interpreting Important Variables

I. Minitab / Computer Printouts

Below is a computer output. You will be expected to use and interpret computer output on the AP Exam. This output is from Minitab, however most computer output looks very similar. We will discuss which numbers you need to know, what they mean, and how to interpret them.



MINITAB OUTPUT FOR BOX 2.1 Analysis of the *trees* dataset: regression

Regression Analysis: VOLUME versus HEIGHT

The regression equation is
VOLUME = - 87.1 + 1.54 HEIGHT

Predictor	Coef	SE Coef	T	P
Constant	-87.12	29.27	-2.98	0.006
HEIGHT	1.5433	0.3839	4.02	0.000

S = 13.40 R-Sq = 35.8% R-Sq (adj) = 33.6%

Analysis of Variance

Constant -----→ -87.12

This is the y-intercept. This is the value of the response variable when the explanatory variable is 0. Check the context of the situation. Often, there can be no such value.
In this case, it is not possible to have a volume that is negative nor is it possible to have a height of zero.

Height/Slope ---→ 1.5433

This is the coefficient of the explanatory variable, thus it is the slope.
This entire line of numbers deals with regression for slope.
For each increase in height of one unit, the volume is expected to increase by ***approximately*** 1.5433 units. (Actual units were not provided)

Prediction equation-----→ $y = -87.1 + 1.54x$ (ie. Least Squares Regression Line)

This is an equation used to make predictions and is based on only one sample.

SE Coef ---------→ 0.3839

This is the standard deviation of the slope. Remember, this data came from only one sample. We would expect the slope to vary a little from sample to sample. Thus,

If we gathered repeated samples, we would expect the slope of the volumes of the trees to vary by approximately 0.3839 units.

S ---------→ 13.40

This is the standard deviation of the residuals. The average amount that the observed values differ from the predicted values is 13.40. The average amount that the observed volumes of trees differ from the predicted volumes is approximately 13.40 units.

AP Statistics – AP Exam Review **Name:_____**

Linear Regression & Computer Output: Interpreting Important Variables

r^2 ----------→ 35.8%

This is the correlation of determination, which is the fraction or proportion of variation in the y values that is explained by the least squares regression of y on x.

About 35.8% in the variation in volume can be explained by the least squares regression of y(volume) on x (height).

$r = \sqrt{r^2} = \sqrt{.358} = .598$

This is the correlation coefficient. It tells you strength and direction of the relationship.

With an r value of .598, there is a weak, positive relationship between height and volume of trees.

T ----------→ 4.02

This is the test statistic which = $test\ statistic = \dfrac{statistic - parameter}{std.dev\ of\ statistic} = \dfrac{b}{SE}$

P---------→ 0

This is the p-value of a Linear Regression t test.

With a p-value of approx.. 0 less than any alpha level (.05, .01), reject the null. There is evidence that there is a relationship between the volume of a tree and its height.

Example 2: Minitab / Computer Printouts

Regression Analysis: Height versus Mother Height

```
The regression equation is
Height  =  24.7 + 0.640 Mother Height
```

(dependent variable, y) (intercept, b_0) (slope, b_1) (independent variable, x)

⟸ the estimated regression equation: $\hat{y} = b_0 + b_1 x$

Predictor	Coef (estimates)	SE Coef (sd of ests.)	T (test statistics)	P (p-values)	
Constant	24.690	8.978	2.75	0.009	IGNORE these values
Mother H	0.6405	0.1394	4.59	0.000	tests H_0: $\beta_1 = 0$, vs. two-tailed altern.

intercept, b_0 — slope, b_1

(the latter is equivalent to testing for linear correlation between x and y)

```
S = 2.973        R-Sq = 35.7%        R-Sq(adj) = 34.0%
```

(standard error of estimate, s_e) (coefficient of linear ***determination***, r^2) (adjusted r^2, used for multiple regression)

* (the coefficient of linear ***correlation*** is the square root of r^2, with the same sign as the slope, b_1)

for a simple linear regression minitab models only conduct two-tailed alternatives

S – 2.973 standard error of the estimate, or the standard deviation of the residuals - actual values versus the predicted values

SE Coef – 0.1394 – standard deviation of the slope

AP Statistics – AP Exam Review **Name:_____**

Linear Regression & Computer Output: Interpreting Important Variables

Practice Minitab / Computer Printouts

1. A sample of men agreed to participate in a study to determine the relationship between several variables including height, weight, waste size, and percent body fat. A scatterplot with percent body fat on the y-axis and waist size (in inches) on the horizontal axis revealed a positive linear association between these variables. Computer output for the regression analysis is given below:

```
Dependent variable is: %BF
R-squared = 67.8%
S = 4.713 with 250-2 = 248 degrees of freedom

Variable     Coefficient        se of coeff        t-ratio      prob
Constant       -42.734             2.717            -15.7      <.0001
Waist            1.70              0.0743            22.9      <.0001
```

(a) Write the equation of the regression line (be sure to use correct notation and define your variables):

(b) Explain/interpret the information provided by R-squared in the context of this problem. Be specific.

(c) Calculate and interpret the correlation coefficient (r).

(d) One of the men who participated in the study had waist size 35 inches and 10% body fat. Calculate the residual associated with the point for this individual.

AP Statistics – AP Exam Review **Name:_____**

Linear Regression & Computer Output: Interpreting Important Variables

Determine the LSRL, Standard deviation for slope, correlation coefficient and the standard error of the residuals for each:

#2.

MINITAB OUTPUT FOR BOX 2.1 Analysis of the *trees* dataset: regression

Regression Analysis: VOLUME versus HEIGHT

The regression equation is

Predictor	Coef	SE Coef	T	P
Constant	-87.12	29.27	-2.98	0.006
HEIGHT	1.5433	0.3839	4.02	0.000

S = 13.40 R-Sq = 35.8% R-Sq (adj) = 33.6%

#3.

MINITAB OUTPUT FOR BOX 2.3 Analysis of the *scores* dataset: regression

Regression Analysis: MATHS versus ESSAYS

The regression equation is

Predictor	Coef	SE Coef	T	P
Constant	27.57	22.26	1.24	0.251
ESSAYS	0.6548	0.3154	2.08	0.072

S = 9.140 R-Sq = 35.0% R-Sq(adj) = 26.9%

Linear Regression & Computer Output: Interpreting Important Variables

II. More Practice with Linear Regression and Residual Plots

4 . Fast food is often considered unhealthy because much fast food is high in fat and calories. The fat and calorie content for a sample of 5 fast-food burgers is provided below.

Fat(g)	Calories
31	580
35	590
39	640
39	680
43	660

a) Identify the explanatory and the response variables:

b) Use the calculator to make a scatter plot of these ordered pairs. Sketch the scatter plot here.

c) What information does the scatter plot provide? That is, use the scatter plot to describe the relationship between the fat grams and calories in a fast food burger.

d) Find the following summary statistics for this data:

$\bar{x},\ \bar{y},\ s_x,\ s_y$

e) Now use your calculator to record the following statistics and to find the equation of the least squares line. Record the equation and use it for the remaining computations.

$a,\ b,\ r^2,\ r,\ y$

f) Examine a graph of the least squares line superimposed on your scatter plot.

Stat > calc > 8:linreg(a+bx) >L1, L2, Y
To get the Y to show up: Vars > Y Vars> 1:Function > 1: Y1

This will graph the LSRL along with your scatterplot. If you go to the Y= screen, you will now see the equation for the LSRL

g) Does the line appear to be good model for the data?

h) What is the value of your slope? What information does it provide? Be specific.

i) How many calories would you predict a burger with 20 fat grams has?

j) Calculate the residual for 35 fat grams.

k) Calculate the value of r^2. What information does it provide? Be specific.

l) What is the value of r ? What does it tell you in this situation?

m) Make a residual plot on your calculator. Be sure to label both axes with words and a "friendly" scale.

n) Based on this residual plot, do you think the least squares line is a good model for this data

Linear Regression & Computer Output: Interpreting Important Variables

5. Becky's parents have kept records of her height since she was born. The data set consists of Becky's age in months and her height in centimeters. The summary statistics for the data are provided below:

Mean age: 44 months	std. dev. age: 8.5 months
Mean ht: 82 cm	std. dev. ht: 4.1 cm

The correlation between age and height is .86.

(a) Find the equation of the least squares line that you would use to predict Becky's height from her age. **Show all work.**

(b) What real-world information does the slope provide? Be specific!

(c) Suppose height had been measured in inches rather than in centimeters. What would be the correlation between age and height in inches? Note: 1 inch = 2.54 cm